下载APP

13 | 大促容量保障体系建设：怎样做好大促活动的容量保障工作（下）

2021-06-09 吴骏龙

《容量保障核心技术与实战》 课程介绍 >

讲述：吴骏龙
时长 12:49 大小 11.74M

你好，我是吴骏龙。

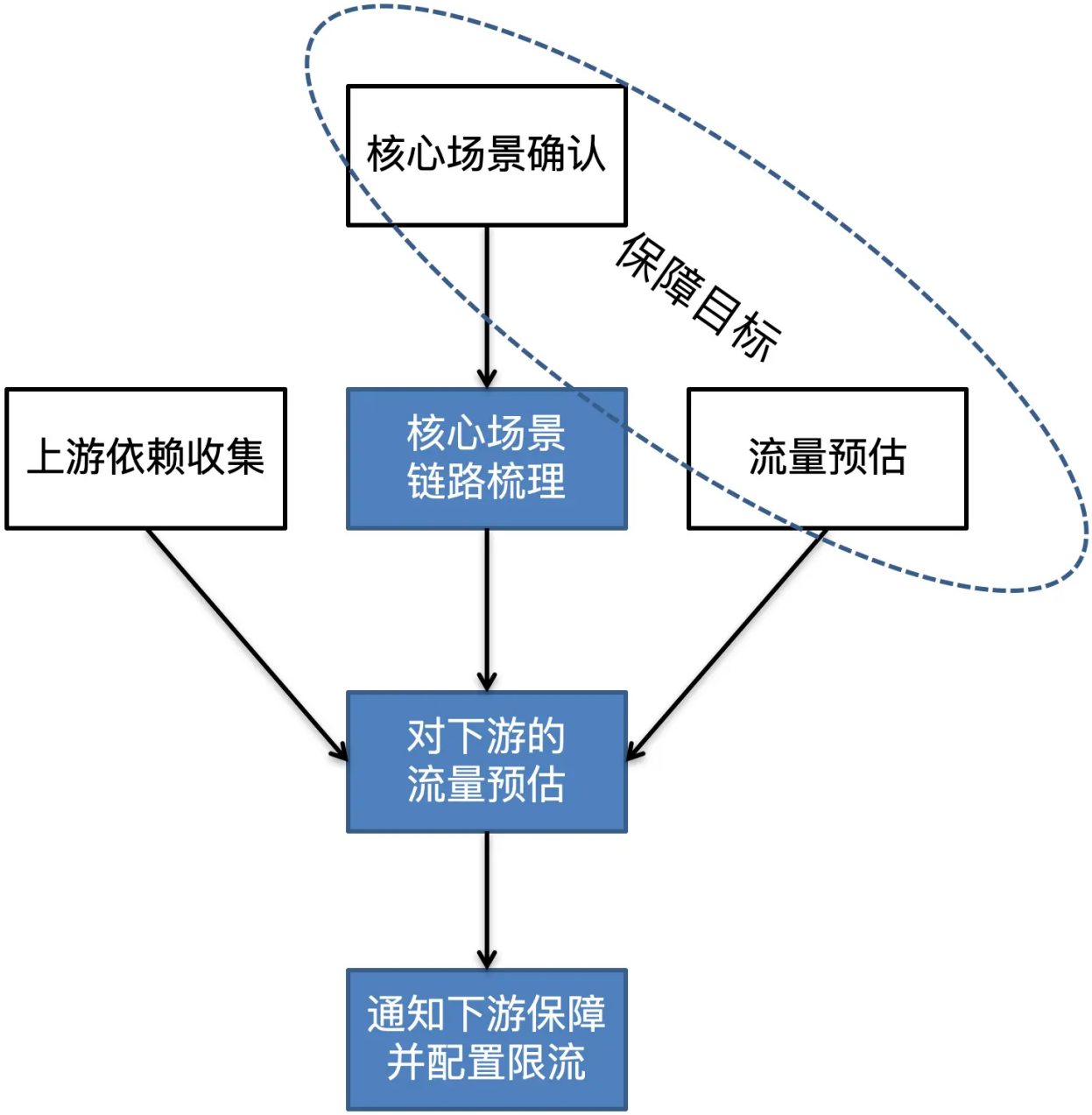
在上一讲中，我提到了大促活动在业务场景和业务指标中的不同特点，并结合本地生活的双 11 大促活动案例做了讲解。同时，基于这些特点，我为你构建了一个大促容量保障的框架，并做了一定的展开，以便你能应用到具体工作中去填空。

上一讲主要涉及大促容量保障的两项准备工作：重点链路梳理和服务架构治理，这一讲我们步入大促容量保障的进行时，看一看大促容量保障的三项重点步骤：**大促流量预估**、**大促容量测试**和**大促容量保障预案**。最后，我会为你总结一下整个大促容量保障的体系大图。

大促流量预估

大促流量预估是实施容量测试的前提，在“容量保障的目标”一讲中，我有谈到的一个服务流量扩散比计算的案例，这就是大促场景流量预估的一个很常见的方法。

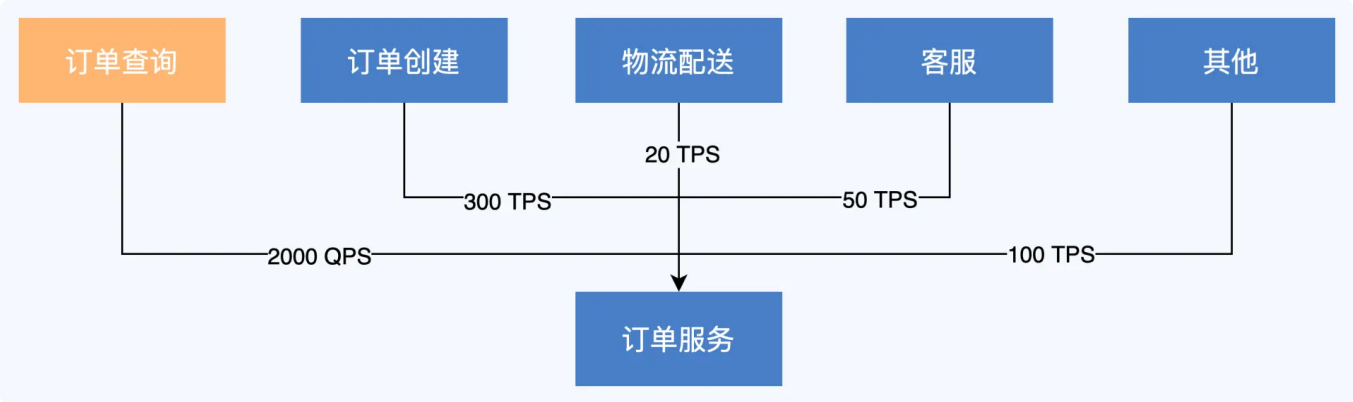
具体再总结一下重点，如下图所示，在梳理出核心场景链路后，对各服务进行流量预估时，特别要注意对下游服务的流量预估，一方面需要从**核心场景**入手计算，另一方面还要考虑到**上游其他服务的依赖**。



这里提供一个不错的实践，也是我在双 11 容量保障工作中沉淀的一个做法，称之为**上游负责制（又称入口服务负责制）**。简单说，所有依赖方应负责预估对被依赖方的调用量，并

及时通知下游进行保障。对于调用量很高，且非常关键的调用方，下游服务甚至可以切割出一个物理隔离的集群，专门为这个调用方服务，以降低容量相互影响的可能性。

我以双 11 活动的订单服务为例，订单服务是非常核心的服务，也是很多其他关键服务的下游，调用量非常大。如下图所示，有大量服务都会调用订单服务执行逻辑，这里我们屏蔽掉调用方式，只从调用量的角度出发，根据“上游负责制”，每个上游服务必须预估对订单服务的调用量，并确保超出时自身先进行限流，或者降级。



在订单服务得到这些信息后，也要做防御性容量保障，不能假定这些上游都按套路出牌。还是按照上图的场景来讲，我们拆分出两个服务：订单查询和订单创建，分别针对订单服务的读操作和写操作，提供对接用户和商户的功能封装，这是典型的读写分离的架构思路。

我们发现，订单查询对订单服务的调用量最大，而且订单查询是非常重要的服务，直接面对用户和商户，为了避免风险，对订单服务来说，就可以切分出一个独立的物理集群，专门为订单查询提供服务。

这样做的好处是，一旦订单查询对订单服务的调用量出现预料外的峰值，导致订单服务出现容量瓶颈，最差也不会影响到其他的调用方。当然，对于后续的容量测试工作，两个集群的容量也需要分别进行验证。

除了上面谈到的案例，我再教给你一些关于大促流量预估的诀窍：

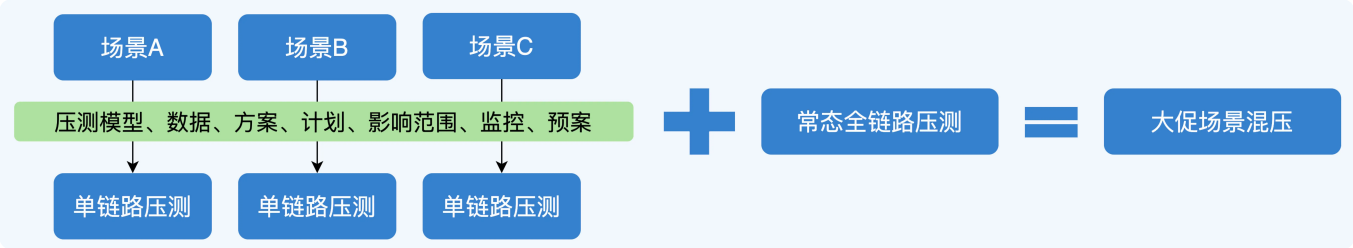
1. **大促期间的通用场景流量预估**：这些场景不是大促活动所特有的，可能是日常业务场景，或者是之前已经举办过的活动场景。针对通用场景，我们可以依据常态的流量情况，结合业务指标加上一定的涨幅就可以了。需要注意的是一定要对比清楚场景，勿简单相加。

2. **二八法则**：流量的分布不是均匀的，如果大促场景没有明确限定时段，**可以预期 80% 的流量会发生在 20% 的时间内**，并依此评估峰值 TPS。
3. **竞品对比**：有些大促场景是全新制定的，在公司内没有任何历史经验可以借鉴，也没有靠谱的业务指标供转化。这时候可以了解一下竞对公司是否有进行过类似的活动，通过收集公开数据进行横向对比，也不失为一种流量预估的手段。

大促容量测试

完成大促流量预估后，我们就有了验收标准，下面进入大促容量测试工作。由于大促活动有其特定的业务场景形态，除了常态的全链路测试工作外，针对大促场景需要进行多次单链路压测，**每次单链路压测都包含压测模型构建、压测数据准备、压测方案和计划准备、影响范围预估，以及监控和预案等步骤。**

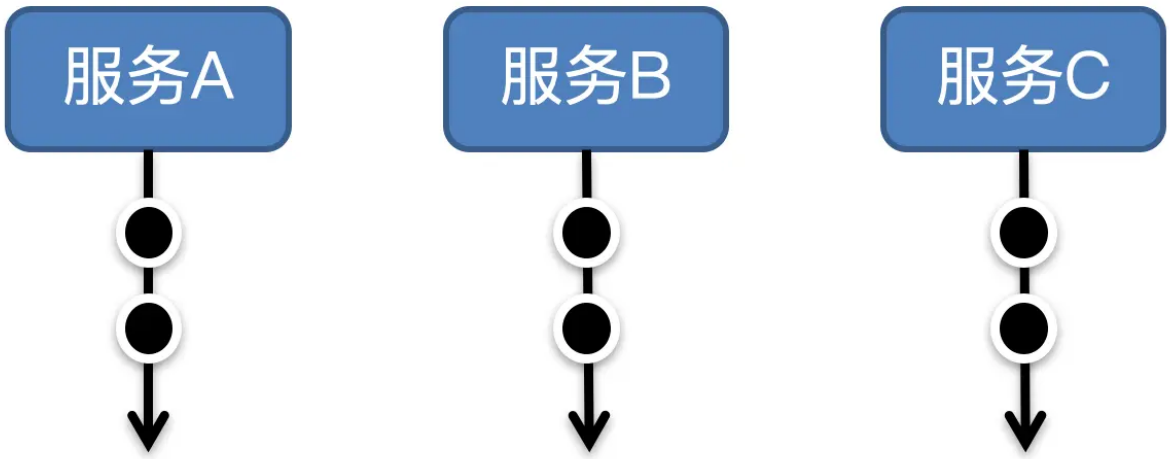
最重要的部分来了，在单链路压测通过后，需要再和全链路压测叠加进行混压，方能确保全局容量不出问题。 虽说混压的成本比较高，但它是必不可少的，单压没有问题只能说明局部最优，无法推导出全局最优。



一个案例：单压满足容量目标，但混压却都不达标？

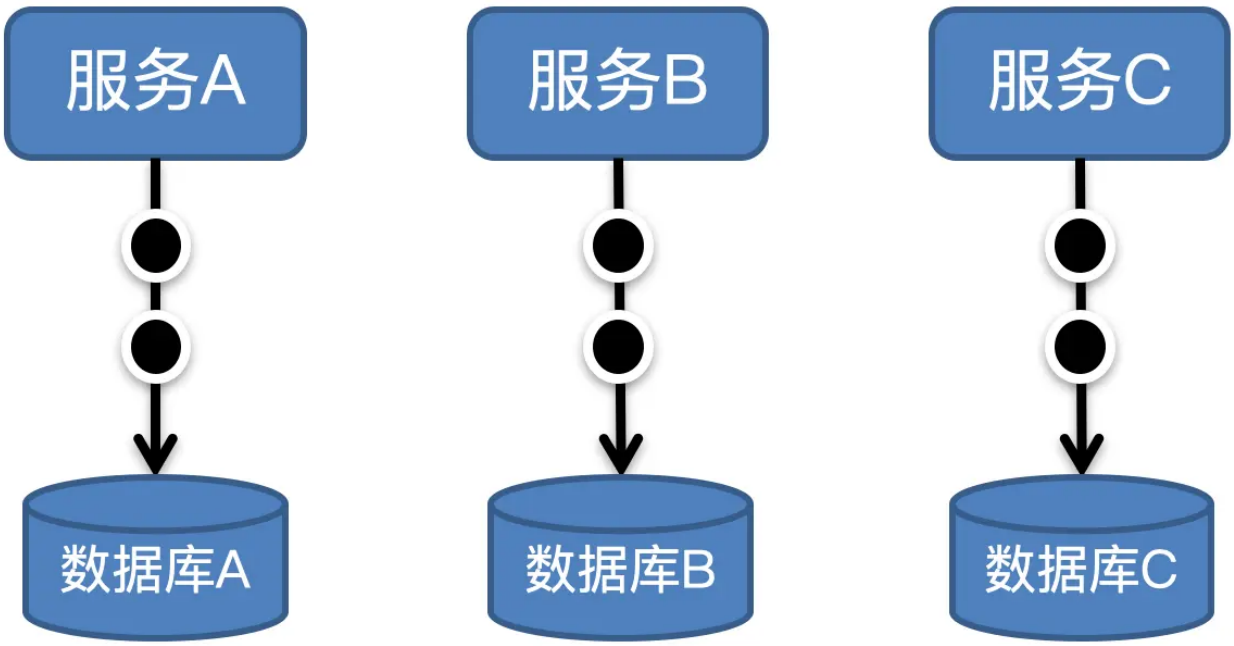
为了说明这个问题，我们来看一个例子。在双 11 大促场景中，某大促活动涉及三个营销服务，我们分别对这三个服务进行单压，都满足容量目标，但在叠加混压时，全部都不达标，这是为什么呢？

我们先检查了一下这些服务的部署方式和调用关系，发现这些服务都是独立部署的，相互之间也没有直接调用关系，从服务链路层面似乎看不出什么问题。



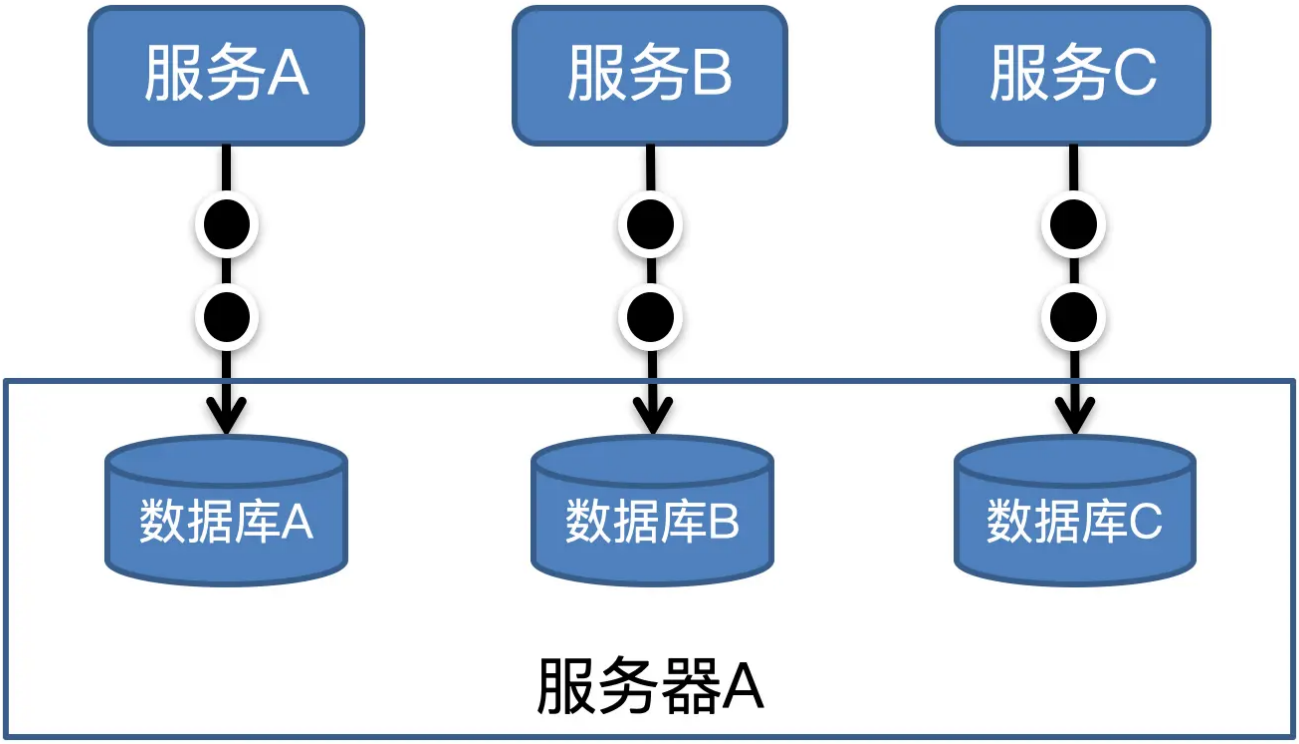
于是，我们顺着链路向下挖掘，一直到最底下的数据库，发现了一些端倪。在单压时，各服务所涉及的数据库读写性能基本都在 10ms 这个级别，性能没有什么问题，但在混压时，这些数据库的读写性能都不约而同的集体变差了，上涨了近 10 倍。

这个现象非常奇怪，对这些数据库的调用除了上述 3 个营销服务外，在混压期间没有其他调用方，相互之间也没有交叉调用的情况，似乎不太会存在影响。



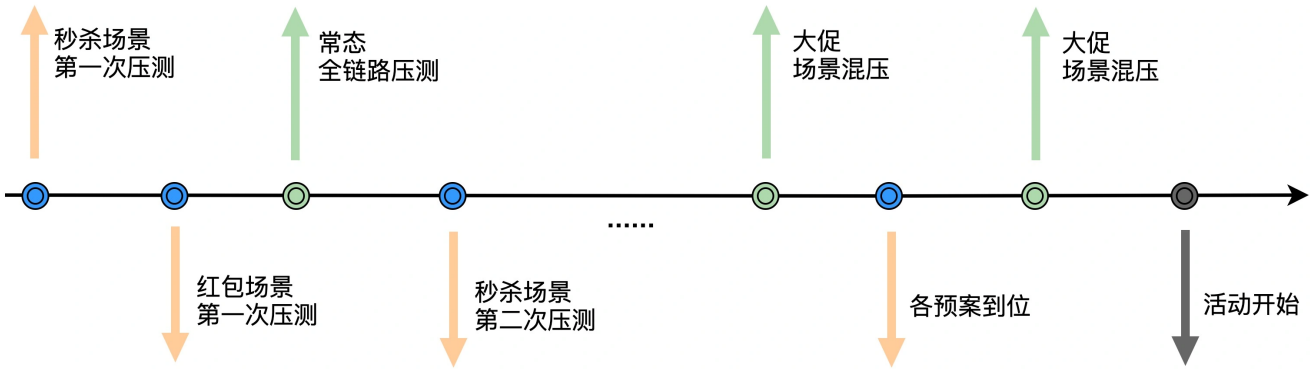
但不管怎么说，既然我们发现混压时主要的瓶颈在数据库上，必须拉上了 DBA（数据库管理员）一起排查，DBA 在观察到数据库服务器的高负载情况后，很快得到了结论。原来是

这三个服务底层的数据库节点恰好是混部在一台服务器上的，结果就是，在混压时数据库无法承载各自单压时的叠加压力，造成容量不足。



在双 11 活动场景下，这三个营销服务肯定是同时处于高负载下的，因此只有通过混压的方式才能还原真实场景，发现这类问题。这个例子虽然经过了一定的简化，但已经充分体现了混压的重要性，哪怕是一些表现上看起来没有什么关联的服务，在底层也有可能会相互牵扯，影响各自的容量表现。**因此，为准确评估双 11 期间的真实容量情况，必须进行混压工作**。

从时间维度来说，可以在大促尚未开始的早期，就安排常态全链路压测，确保常态容量安全。大促场景的单链路压测可以穿插在大促准备期，根据服务上线的计划安排。在大促开始前一段时间会提前封版，确保所有服务不再变更后，密集组织若干次混压，保证全局容量安全。



大促容量保障预案

任何事前的容量保障预防措施都不可能完全规避问题，不能“把鸡蛋放在一个篮子里”，一旦在大促活动期间不幸还是发生容量问题，我们也要有相应的预案能够减少和避免损失。

在“容量治理”那一讲中我们已经谈到了一些常用手段，如扩容、限流、降级等。光有这些手段还是不够的，我们还需要将其通过预案的形式固化下来，真正落到实战中去。

预案是生活和工作中很常见的概念，下面这些描述都属于预案。

高温红色预警 停止户外露天作业	某服务CPU利用率>80%且趋势仍在上升 紧急扩容	某服务突发故障且短时间内无法恢复 服务降级

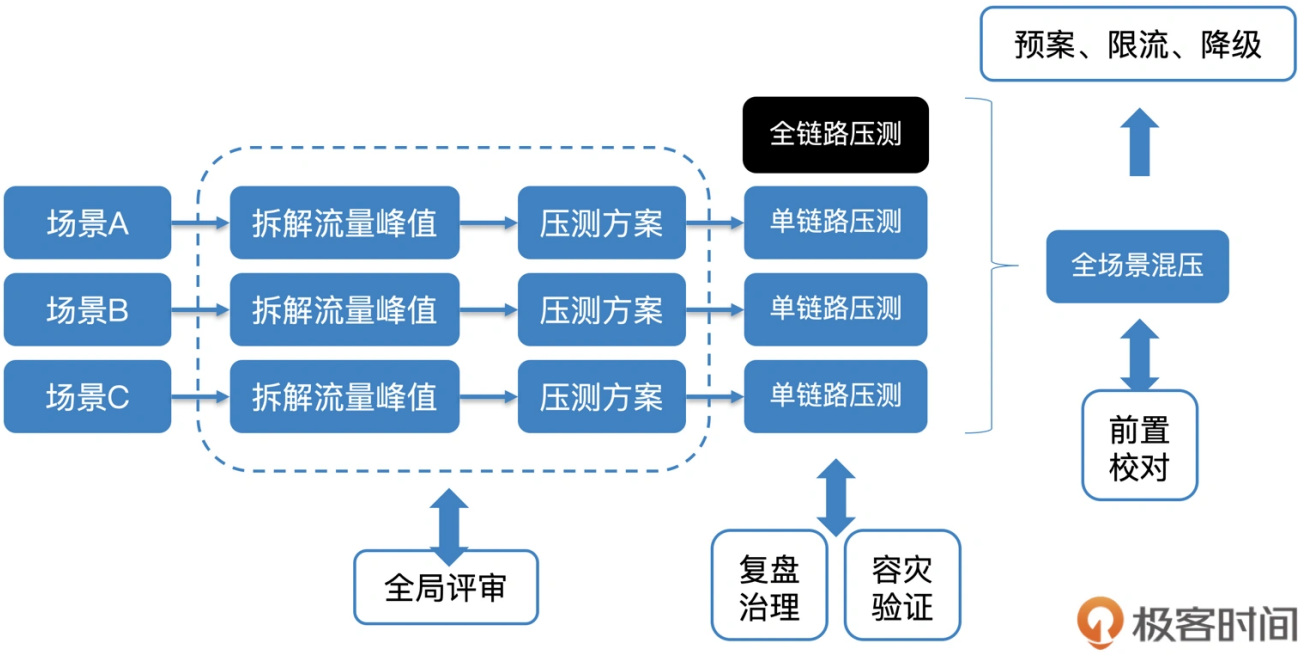
预案编写应非常具体，避免歧义，还要落实到具体的人。我国对应急预案制定有国家标准，在互联网行业我们可以简化一下，将预案组织为：**负责人、服务对象、执行条件、预案内容与操作步骤、影响面等五大要素。**下面是一个双 11 活动期间我们的预案实例，张三所负责的天气服务基于的是第三方天气供应商提供的数据，如果供应商服务出现异常，可以操作降级，效果是不显示天气，用户无法获取准确的天气信息。

预案五要素：负责人、服务对象、执行条件、预案内容与操作步骤、影响面				
张三	天气服务	第三方天气供应商异常	降级，不显示天气	天气情况不准确

完善预案覆盖率是保证业务快速恢复的有效手段，作为主动型的容量保障手段，我们需要保障预案是可执行的，人员能够熟练操作，且影响范围周知。做到这一点的落脚点是**进行演练**，而且是频繁的演练，通过演练提升人员操作熟练度，减少误判，防止预案腐化，这项工作一般会穿插在双 11 活动的准备期进行。

大促容量保障体系

到这里，大促容量保障进行时的三项重点工作就介绍完了。在这一讲的最后，基于上述大促活动的容量保障各项工作，我们将其抽象一层，得到大促容量保障的体系建设大图，你可以直接以它为蓝本，设计你的大促容量保障方案。



这张图里的不少环节，是我摸爬滚打后总结出来的，我也将这些经验与你分享一下。

首先，我们时常会发现一些单场景容量评估存在疏漏的情况，比如遗漏一条链路没有考虑到，或是调用量计算错误，这一类疏漏有时候不一定是人员的不细心造成的，而是因为缺乏足够的信息。为规避这一问题，我们在各个大促场景的容量预估完成后，组织了一次**全局评审**，全局评审的目的是拉齐各方之间的信息，查遗补漏，确实能发现一些单场景梳理缺失的信息。

其次，在一个超大规模技术团队内组织全场景混压也是很令人头疼的工作，需要协调各个团队的技术人员，协调测试数据和资源，还要制定可操作性强的流程，传统的文档说明和宣讲都很难保证每个人都知晓了整体流程。

为此，我们在全场景混压前 1 小时，安排了一次**前置校对**的环节，其实就是一个简单的流程通气会，会议的内容是串讲压测流程，并现场答疑，相当于是一次彩排的过程，彩排结束后，立刻进入实战。至于测试数据和资源协调工作，化整为零在线下逐业务团队异步完成，在彩排前可以先小流量进行调试，确保脚本和数据无误，资源足够。

如果你所在公司的技术团队很大，不妨也可以尝试这种做法。

最后，人的认知是螺旋式发展的，在大促活动完成后，应该及时**组织复盘**，总结经验，将不足之处尽可能多的暴露出来，以免将来再犯。

总结

这一讲，我主要谈到了大促容量保障的三项重点工作：大促流量预估、大促容量测试和大促容量保障预案，并进行了一定的体系化归纳。

相比常态流量，大促流量预估的难度较大，尤其是新上线的服务。对此，我给出了一些实际工作中总结的方法和技巧，以及比较有效的行为准则，如上游负责制、二八法则等。

大促容量测试的核心是混压，组织协调是关键，可以通过拆分出一系列的单链路压测，先行识别容量瓶颈，再通过混压的方式进行验收，达到验证全局容量的目的。

未雨绸缪早当先，居安思危谋长远。对于大促这样的一次性活动形态，除了要制定完善的容量治理手段，还要通过预案的方式进行固化，并通过演练避免无效的预案，保证活动期间一旦出现容量问题，能够快速止损。

最后，我将大促容量保障工作抽象出一张体系大图，并讲解了实际遇到的一些困境和解决方法，希望对你有所帮助。

课后讨论

请你根据今天的所学所得，试着编写一些容量保障的预案，并说一说如何组织演练？欢迎与我交流你的想法。

分享给需要的人，Ta订阅后你可得 16 元现金奖励

赞 1 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 12 | 大促容量保障体系建设：怎样做好大促活动的容量保障工作（上）

下一篇 14 | 容量保障组织建设：容量保障需要什么样的团队？

精选留言

写留言

> 由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。